Technical Research Note 187

AD

IMPACT OF FEEDBACK ON ACCURACY OF CONFIDENCE LEVELS ASSIGNED BY INTERPRETERS

by James A. Thomas and Robert Sadacca

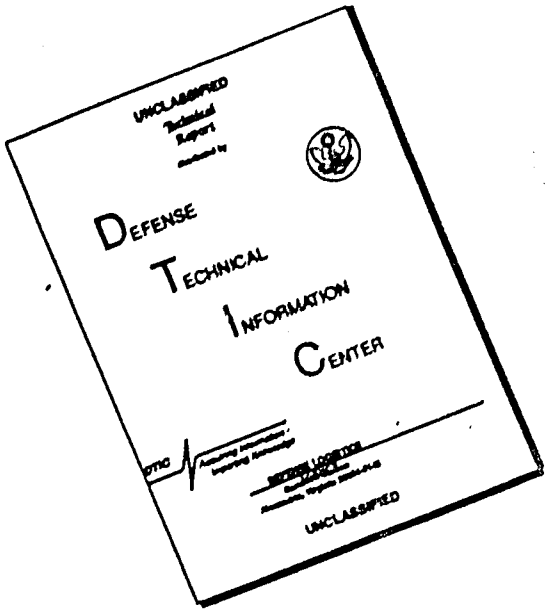
Support Systems Research Division

JUNE 1967

U. S. Army
Behavioral Science Research Laboratory

27

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

Technical Research Note 187

AD

IMPACT OF FEEDBACK ON ACCURACY OF CONFIDENCE LEVELS ASSIGNED BY INTERPRETERS

by James A. Thomas and Robert Sadacca

SUPPORT SYSTEMS RESEARCH DIVISION
Joseph Zeidner, Chief

U. S. ARMY BEHAVIORAL SCIENCE RESEARCH LABORATORY

Office, Chief Research and Development
Department of the Army

Washington, D. C. 20315

June 1967

Army Project Number
2J620901A721

Man-Computer Functions b-00

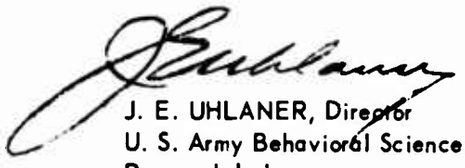
BESRL Technical Research Reports and Technical Research Notes are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The Surveillance Systems Project has as its objective the development of scientific research data bearing on the extraction of information from imagery and the products of other sensors, and the efficient storage, retrieval, and transmission of information within an advanced computerized image interpretation facility. Research results are used in future systems design and in the development of enhanced techniques and procedures for all phases of the image interpretation process within the data reduction facility.

The MAN-COMPUTER FUNCTIONS Task is one of four research Tasks established in the Support Systems Research Division of the Behavioral Science Research Laboratory to concentrate on operational segments of the surveillance system. One major effort of the task is devoted to the transfer and control of intra-system information. The objective is to develop techniques whereby the computer can increase the effectiveness of interpretation by supplementing the decision processes of the interpreter, performing his routine calculations, and evaluating the accuracy and completeness of his interpretations. The present study dealt with the utility of feedback presented under simulated computerized conditions in improving the performance of interpreters in judging the value of their own reports.

The entire research program is responsive to objectives of RDT&E Project 2J620901A721, "Surveillance Systems: Ground Surveillance and Target Acquisition Interpreter Techniques". FY 1966 Work Program.



J. E. UHLANER, Director
U. S. Army Behavioral Science
Research Laboratory

IMPACT OF FEEDBACK ON ACCURACY OF CONFIDENCE LEVELS ASSIGNED BY INTERPRETERS

BRIEF

Requirement:

BESRL research has indicated that interpreters tend to have more confidence in the accuracy of the information they extract from imagery than is warranted. An effective method of improving the accuracy with which interpreters evaluate their identifications must be found before their evaluations can be used to full advantage operationally. The present study explored the effect on subsequent performance of giving interpreters knowledge-of-results practice in rating the accuracy of their identifications.

Procedure:

A different technique of providing feedback was employed with each of three experimental groups of interpreters. A control group received no feedback. The 15 interpreters in each group reported on three sets of imagery, one set on each of three successive days, the same order of presentation being followed in all groups. In the three experimental groups, feedback based on the previous day's performance was provided at the start of the sessions on the second and third days. In Technique A, the interpreter was given a summary sheet containing a distribution of his previous confidence ratings and stating whether his ratings were overestimates or underestimates of the accuracy of his identifications. In Technique B, in addition to the summary sheet, the interpreter was given his scored answer sheets from the previous session, along with the imagery he had interpreted. In Technique C, in addition to the information given in Technique A, the interpreter was given a distribution of ratings and accuracy scores purported to have been made by several previous classes of interpreters.

Findings:

Results supported the previous findings that interpreters do not as a rule make dependable evaluations of their identifications. However, the confidence ratings made by interpreters whose identifications were generally more accurate and complete were more precise than those made by interpreters in a low performance group.

Feedback techniques A and C, in which interpreters were given only data on previous rating performance--their own (A) and their own plus that of other classes (C)--resulted in somewhat more accurate expressions of confidence than did Technique B in which interpreters were given their own corrected reports and the imagery they had interpreted in a previous session. The confidence ratings of the control group, which had received no feedback, were the least precise.

Utilization of Findings:

The improvement noted with two feedback methods indicates that the accuracy of interpreters' confidence ratings can be increased by practice in applying a knowledge-of-results frame of reference. The improved confidence ratings, however, were still generally inaccurate. Evidently, more than two practice sessions are needed to enable the interpreter to reach an operationally useful level of accuracy in evaluating the information he provides.

IMPACT OF FEEDBACK ON ACCURACY OF CONFIDENCE LEVELS ASSIGNED BY INTERPRETERS

CONTENTS

	Page
INTRODUCTION	1
OBJECTIVES OF THE STUDY	2
EXPERIMENTAL PROCEDURES	2
Research Design	2
Performance Measures	2
Subjects	3
Confidence Judgments	3
Feedback Techniques	3
Variables	4
ANALYSIS AND RESULTS	5
Overall Accuracy of Confidence Ratings	5
Effect of Feedback on Confidence Inaccuracy Scores	6
Effect of Feedback on Mean Completeness and Accuracy Scores	8
IMPLICATIONS	8
DISTRIBUTION	9
APPENDIXES	11
DD Form 1473 (Document Control Data - R&D)	21

TABLES

Table 1.	Mean confidence inaccuracy scores for high and low performance groups	6
2.	Mean confidence inaccuracy scores	7
3.	Mean completeness scores	7
4.	Mean accuracy scores	8

FIGURES

Figure 1.	Percentages of correct identifications made at various confidence levels by all interpreters on Set 1 imagery	5

IMPACT OF FEEDBACK ON ACCURACY OF CONFIDENCE LEVELS ASSIGNED BY INTERPRETERS

In aerial surveillance systems, image interpreters are a vital link between the aerial platforms with their cameras and other sensor devices and the intelligence consumer. The task of the image interpreter is to extract accurate intelligence information from the surveillance imagery.

Accurate reporting means more than making identifications of enemy targets and positions. It means presenting the information so that intelligence consumers have a basis for judging the reliability of the information. To this end, it is common practice in image interpretation facilities for an interpreter to qualify the information in his report as "positive", "probable", or "possible". In most military situations, in fact, the full intelligence potential of an image is not exploited unless the "possibles" and "probables" as well as the "positives" are reported. The image interpreter's confidence in his identifications can then be weighed by intelligence consumers who utilize the reported information in making decisions.

The qualitative categories serve to indicate the interpreter's judgment only in a very general way. Too, the words themselves are somewhat ambiguous. They may have different connotations for different image interpreters--and for different intelligence users. To counter the likelihood of ambiguity, BESRL has introduced and used in all recent image interpretation research a quantitative scale ranging from zero to 100. It was felt that using a quantitative scale to indicate degree of confidence would reduce ambiguity, increase the range of confidence values that could be expressed, and allow more rapid handling of confidence information by automatic data processing equipment.

Subsequent studies, however, showed that image interpreters frequently tend to have more confidence in their identifications than is warranted by the accuracy of the identifications. The accuracy rate of some interpreters for identifications about which they expressed high confidence has been found on some BESRL performance measures to be less than 50 percent[1]. An effective method of improving the accuracy with which confidence ratings are made must be found before such ratings can be used more fully operationally. The present study explored the effect on subsequent judgment performance of giving interpreters practice sessions designed to improve their accuracy in assessing their confidence in their own reports.

---

[1] Cadacca, R, Martinek, H., and Schwartz, A. Image Interpretation Task--Status Report. BESRL Technical Research Report 1129, June 1962.

OBJECTIVES OF THE STUDY

The primary objective of the present study was to determine whether practice in making confidence ratings can improve the accuracy of the ratings. The method consisted of providing the interpreter with feedback and then comparing his stated confidence levels with the scored accuracy of his identifications. Three different techniques of providing feedback were tried out.

A second objective was to determine whether practice in utilizing feedback affects the accuracy with which the interpreters identify targets or the completeness of their interpretation of an image. It was conceivable that emphasis on making accurate confidence ratings would influence the number of correct and incorrect target identifications made by interpreters.

EXPERIMENTAL PROCEDURES

Research Design

Three experimental groups and one control group, each containing 15 subjects, were formed. A different technique of providing feedback was employed with each experimental group. All interpreter subjects examined three sets of performance measures in the same order, one set on each of three successive days. Confidence feedback based on the previous day's performance was provided the interpreters at the beginning of the second and third sessions. Subjects in the control group received no feedback.

Performance Measures

The performance measures consisted of three sets of conventional black and white photographs typical of the operational imagery which confronts the image interpreter. As in the operational situation, interpreters were provided with maps, sortie plot overlays, and standard references and photo keys. They were also given situation sheets showing the number of photographs in the performance measures, the scale of the photos, the intelligence information requested, and the battlefield situation at the time the photos were obtained. The situation sheets were read aloud to the interpreters before they began to examine the photographs (See sample situation sheet in Appendix B).

The interpreters were asked to detect and identify objects of military significance such as wheeled vehicles, artillery, armor, and fortifications. They marked directly on the photographs the objects they located and then recorded identifications of the objects on special answer sheets, using only the descriptive terminology provided in the Target List (Appendix B).

Performance on Set 1 (T-4, T-14, and T-22)[2] was used to match the subjects in the experimental groups and to provide confidence feedback. Performance on Set 2 (T-8 and T-10) was used to measure the effect of the feedback and to provide confidence estimates for feedback immediately before administration of the third set. Set 3 (T-6 and T-3) was used to measure the effect of previous feedback.

Subjects

Sixty image interpreter trainees about to graduate from the U. S. Army Intelligence School at Fort Holabird, Maryland were the subjects. They were divided into four groups of 15 subjects each, matched on the basis of their performance in assigning confidence ratings to identifications made on the three performance measures of Set 1.

Confidence Judgments

Each interpreter was asked to state the degree of confidence he felt in each target identification he made, using a quantitative scale ranging from zero to 100%. Instructions for using the scale specified that 100% of the identifications to which the interpreter assigned a confidence rating of 100 should be correct, 80% of the identifications with a confidence rating of 80 should be correct, and so forth. The interpreter was thus asked to rate directly the probability that a given identification was correct. Interpreters were cautioned not to over- or underestimate their confidence in an identification (See Appendix A for complete instructions on recording confidence).

Feedback Techniques

The basic feedback principle employed was to present an image interpreter with the accuracy rate he had achieved in identifying objects in the imagery for each level of confidence, along with an indication of the amount of his over- or underconfidence. Presumably, if an image interpreter is shown through feedback that he is consistently overconfident in his ratings, he will revise his judgmental processes and make more realistic confidence ratings.

Three feedback techniques were employed. In Technique A, the feedback consisted of presenting each subject with a summary sheet containing a distribution of his confidence ratings, the percentage of correctly identified targets for each confidence interval, and an indication as to

---

[2] Designations refer to performance measures in the BESRL imagery library.

whether his ratings were overestimates or underestimates (See sample summary sheet in Appendix B). In Technique B, each subject was presented with the scored answer sheets and imagery from his previous performance measure in addition to the summary sheet described above. The subject was thereby able to review his identifications and determine where he had made his errors as well as to study how accurately he had assigned his confidence values. Technique C employed group-oriented feedback. Each subject was given the summary sheet described above. In addition, he was given a distribution of confidence ratings and accuracy rates purported to have been made by several previous classes of image interpreters (Appendix B). These figures showed that good agreement between confidence ratings and accuracy rates had been achieved in previous classes, but that on the average the interpreters had been somewhat overconfident. If Festinger's theory of cognitive dissonance[3] applies here, the subject would presumably experience dissonance due to his deviation from the group norm. His attempt to reduce this dissonance would influence the cognitive processes involved in assigning confidence ratings and he would assign more realistic assessments.

Variables

The effects of the feedback techniques were determined for the dependent variables listed below. Values for these variables were summed across each subject's responses to each set of performance measures.

Confidence Inaccuracy Score. A score expressing the degree of inaccuracy of the confidence ratings made by the interpreter, using the formula suggested by Adams and Adams[4]:

$$\frac{\sum \left| p_i - P_i \right| n_i}{\sum n_i}$$

where p_i is the actual percentage of correct identifications made at stated confidence level P_i, and n_i is the number of ratings made at confidence level P_i. For this variable, larger scores indicate less accuracy.

Completeness Score. The number of right identifications divided by the total number of targets in the imagery.

---

[3] Festinger, Leon. Theory of cognitive dissonance. Evanston, Illinois: Row. 1957.

[4] Adams, P. A. and Adams, J. K. Realism of confidence judgments. Psychological Review. 68, 33-35, 1961.

Accuracy Score. The number of right identifications divided by the total number of identifications made by the interpreter.

ANALYSIS AND RESULTS

Overall Accuracy of Confidence Ratings

Based on the identifications made by all 60 subjects to the Set 1 performance measures, the percentages of correct identifications were plotted by confidence level (Figure 1). Since the confidence scale was defined in terms of probability of being correct, the percentage correct for any given confidence level should ideally equal the confidence level. This ideal relationship is shown by the straight line in Figure 1. It is readily apparent that the interpreters in the sample were generally overconfident. For identifications felt to have low probability of being correct, however, they tended to be less confident than was warranted.

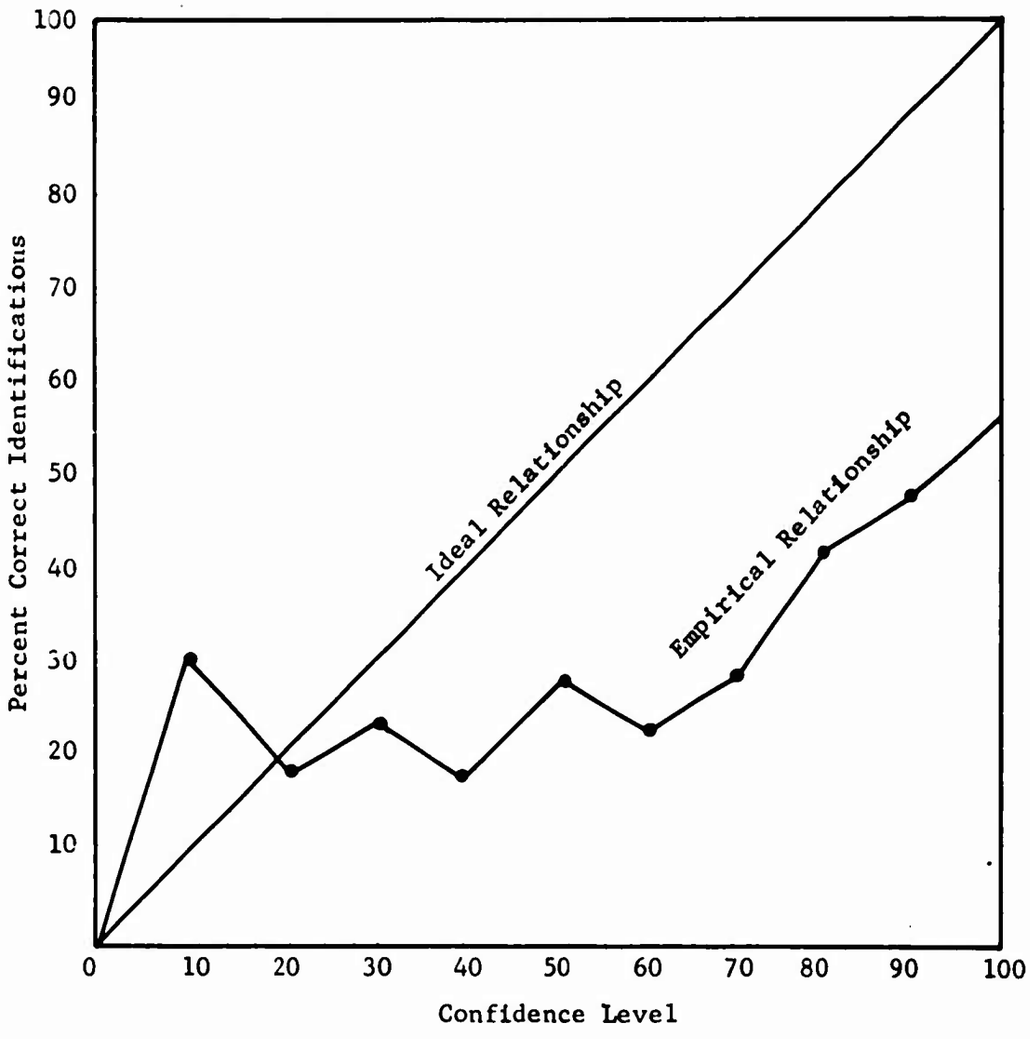


Figure 1. Percentages of correct identifications made at various confidence levels by all interpreters on Set 1 imagery

To determine whether accuracy of confidence ratings was related to interpreter performance, the total sample was divided into high performance and low performance groups on the basis of total right and wrong scores made on the performance measures in Set 1. As shown in Table 1, the confidence ratings made by the interpreters in the high performance group were generally more accurate than the ratings made by those in the low performance group. Further analysis indicated that the low performance group was more overconfident in making their ratings.

Table 1

MEAN CONFIDENCE INACCURACY SCORES
FOR HIGH AND LOW PERFORMANCE GROUPS

	High Group (N = 30)	Low Group (N = 30)
Mean[a]	35.33	43.23
Sigma	6.62	8.09

[a]Means Significantly Different ($p < .01$).

Effect of Feedback on Confidence Inaccuracy Scores

To determine whether the feedback treatments had any effect on the accuracy of the confidence ratings, a two-way analysis of variance (feedback techniques x sets of performance measures) was computed. A repeated-measures design was used for this analysis, since each subject was administered the same sets of performance measures, thereby serving as his own control. The analysis of variance (Table C-1 of the Appendix) produced a significant F-ratio for both the feedback technique and performance measure main effects ($p < .05$).

As shown in Table 2, the mean confidence inaccuracy score was significantly lower for feedback techniques A and C. In addition to the control group, the group using Technique B in which the interpreter reviewed his previous identifications and examined his scores was least effective. Interpreters receiving Technique B feedback may have paid less attention to the summary sheets than the A and C groups in which the summary sheet was the major element in the feedback. The group norms presented to the group employing Technique C may have served to highlight inaccuracies--mean inaccuracy scores were slightly lower for Technique C than for Technique A, where only the summary sheets were used.

The mean confidence inaccuracy score was significantly higher for Set 3 performance measures than for Set 2. Set 3 imagery seemed in general more difficult to interpret judging from the scores made by interpreters (Tables 3 and 4).

Table 2

MEAN CONFIDENCE INACCURACY SCORES[a]
(N = 15 using each technique)

Group	Performance Measure Set 1	2	3	Mean (2, 3)
Feedback Technique A	39.1	32.0	38.9	35.5
Feedback Technique B	39.7	39.9	43.7	41.8
Feedback Technique C	39.3	29.8	35.3	32.5
Control Group	39.5	43.2	49.4	46.3
MEAN	39.4	36.23	41.82	

[a]Means significantly different among feedback techniques ($P < .05$) and between sets 2 and 3 ($P < .05$). Set 1 scores were not included in the analysis of variance.

Table 3

MEAN COMPLETENESS SCORES[a]
(N = 15 using each technique)

Group	Performance Measure Set 1	2	3	Mean (2, 3)
Feedback Technique A	.30	.28	.13	.22
Feedback Technique B	.30	.31	.19	.25
Feedback Technique C	.27	.27	.17	.22
Control Group	.27	.25	.15	.20
MEAN	.29	.28	.16	.22

[a]Means significantly different between Sets 2 and 3 ($P < .001$). Set 1 scores were not included in the analysis of variance.

Table 4

MEAN ACCURACY SCORES[a]
(N = 15 using each technique)

Group	Performance Measure Set 1	2	3	Mean (2, 3)
Feedback Technique A	.42	.41	.28	.35
Feedback Technique B	.39	.44	.37	.41
Feedback Technique C	.38	.42	.37	.40
Control Group	.39	.35	.28	.37
MEAN	.39	.40	.32	.36

[a]Means significantly different between Sets 2 and 3 ($P < .01$). Set 1 scores were not included in the analysis of variance.

Effect of Feedback on Mean Completeness and Accuracy Scores

To determine whether the feedback treatments had any effect on general interpretation performance, a similar analysis of variance was computed for the completeness and accuracy scores. The only significant F-ratios were for performance measure sets.

IMPLICATIONS

The primary objective of the study was to determine whether practice directed at improving confidence ratings would increase the accuracy of the ratings. The method consisted of providing the interpreter with feedback information comparing his stated confidence in his identifications with the scored accuracy of the identifications. The secondary objective was to determine whether the feedback practice affected target identification accuracy and completeness.

It was concluded that practice in which the interpreter is provided with feedback information as a frame of reference against which he can assign confidence levels to identifications he is currently making shows promise for improving the accuracy of confidence ratings. However, more than two practice sessions are necessary for the interpreter to reach an operationally acceptable level of accuracy. How many feedback sessions are needed for the interpreter to reach the desired level of accuracy in making confidence ratings remains to be determined.

DISTRIBUTION

U. S. Army Behavioral Science Research Laboratory

DISTRIBUTION LIST

Directorate for Armed Forces I and E
Director, Army Research, OCRD
Deputy Chief of Staff for Personnel
Assistant Chief of Staff for Force Development
Assistant Chief of Staff for Intelligence
Chief of Personnel Operations, DA
CG, U. S. Continental Army Command
CG, U. S. Army Combat Development Command
CO, U. S. Army Enlisted Evaluation Center
Chief of Information, DA
Chief of Chaplains, DA
Assistant Secretary of Defense for Education
CG, Automatic Data Field Systems Command
Comdt., Marine Corps
Director, Human Resources Research Office
Directors of Research, HumRRO Field Divisions
U. S. Army Medical Research Laboratory, Psychology Division
CO and Director, U. S. Naval Training Devices Center
CG, U. S. Army CEDEC
CG, U. S. Army Electronic Proving Ground
OIC, U. S. Naval Medical NP Research Unit
Director, WRAIR, Walter Reed Army Medical Center
Chief, Personnel Research Staff, OP, U. S. Department of Agriculture
USES, BES, U. S. Department of Labor
The Adjutant General's Office, Personnel Services Support Directorate
Chief of Naval Personnel
Office of Naval Research
Special Operations Research Office
Director, National Security Office
Director, Central Intelligence Agency
Chief, Office of Personnel, PHS, Department of Health, Education, and Welfare
Chief, U. S. Army R and D Office (Panama)
Office of the Provost Marshall General
Office of the Surgeon General, DA
CG, U. S. Army Materiel Command
CG, U. S. Army Security Agency
Director of Rsch and Dev., U. S. Army Electronics Command
Head, Psychology Labs., U. S. Army Natick Laboratories
CO, U. S. Army R and D Group (FE)
CG, Aberdeen Proving Ground, Hum Engr Lab
Chief, Bu M and S, Department of the Navy
Director, U. S. Naval Research Laboratory

APPENDIXES

Appendix		Page
A.	INSTRUCTIONS TO SUBJECT INTERPRETERS	13
B.	MATERIALS USED IN THE EXPERIMENT	15
C.	ANALYSIS OF VARIANCE TABLES	19

APPENDIX A. INSTRUCTIONS TO SUBJECT INTERPRETERS

A-1. INSTRUCTIONS FOR RECORDING CONFIDENCE

Your task is to record how confident you are that your identification is correct in the column labeled "Conf." You are to use a scale that runs from 0 to 100, where 100 indicates that you are certain your identification is correct. If you use this scale accurately, all of the identifications for which you indicate 100% confidence should be correct; 80% of the identifications for which you indicate 80% confidence should be correct; 50% of the identifications for which you indicate 50% confidence should be correct, and so forth. You can use 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 to indicate your estimate of the probability that you have made a correct identification. If you believe you can make finer judgments, you may do so, that is, you might want to use 75 or 95 or even 99, depending on your degree of confidence that your identification is correct.

From previous experiments, we have found that an interpreter's statements of confidence in his identifications are very important in evaluating the accuracy of his identifications; so try to be as accurate as possible. Try not to overestimate or underestimate your confidence. After your identifications have been scored, your ability to estimate your confidence accurately will be determined by comparing your stated confidence with the percent of correct identifications you actually made.

A-2. INSTRUCTIONS FOR EXAMINEES

TO BE READ TO GROUPS A AND C AFTER T _____ HAS BEEN DISTRIBUTED BUT BEFORE THE SITUATION SHEET HAS BEEN READ.

You will find an additional information sheet in your packets for T _____. Take this sheet out and study it carefully. You will have three minutes to do this. Do this now. Do not discuss this information with any other member of this class.
AT THE END OF THREE MINUTES, PROCEED WITH THE ADMINISTRATION OF T ____.

TO BE READ TO GROUP B

The performance measures for T _____, including your annotations and your scored answer sheet have been returned. Study this material carefully to determine the kinds of errors you made and to ascertain what led you to make these errors. Also check carefully the confidence ratings which you used to express your confidence in both the wrong and right identifications. In addition, study carefully the additional information sheet contained in your packets. Ask any questions you have about any aspect of the scoring. You will be allowed 20 minutes.

AFTER 20 MINUTES: For the next sets of performance measures, I want you to try to improve the accuracy of your confidence ratings. Keep in mind the errors which you made and which led you to be overconfident in some of your identifications and underconfident in others. BE MORE CAREFUL AND TRY TO BE MORE ACCURATE IN MAKING THESE RATINGS.

TO BE READ TO GROUP D

You will not need any additional instructions. Please sit quietly for the next three minutes. DO NOT LOOK AT THE IMAGERY UNTIL I TELL YOU TO DO SO.

APPENDIX B. MATERIALS USED IN THE EXPERIMENT

B-1. SAMPLE SITUATION SHEET

PERFORMANCE MEASURE: T-4 IMAGE INTERPRETATION TASK

CONTENTS: Situation sheet; Immediate Report Form; Photos 1-3 from Mission R3923C; Map of area (scale 1:50,000); T-4 List of Military Objects; Sortie plot overlay; and Situation overlay.

GENERAL SITUATION:

You are a member of the photo interpretation team assigned to the First ROK Corps which is defending the right flank of the Eighth US Army Front in Korea during 1952 and 1953. The action along the entire front has been limited to small scale probes.

SPECIFIC SITUATION:

In July 1953, a North Korean POW stated that he had traveled through the area which appears on photo number 2, and saw troop activity, vehicles, and construction in the valley and on the ridge.

On 20 July 1953, the 45th TRS flew a spot reconnaissance mission of the area. Photos 1-3 have been plotted, and your team chief has annotated areas of suspected activity on photo 2. The scale of all photos is 1:5,700.

REQUIREMENTS:

Locate and identify all weapons, vehicles, and fortifications in areas A, B, and C using only those names appearing in the T-4 List of Military Objects.

You have 30 minutes to complete this report.

B-2. SAMPLE TARGET LIST

T-4 LIST OF MILITARY OBJECTS

AA (antiaircraft)

Gun (direct fire artillery, not AA)

How (howitzer)

Mortar

AW (automatic weapon, not part of firing trench)

Missile (missile or rockets)

- - - - - - - - - - - - - - - - -

Car (civilian type)

Light truck (3/4-ton and less)

Truck (larger than 3/4-ton truck, cargo and personnel)

Trailer truck (trailer and tractor)

Construction W (road scrapers, rollers, and other wheeled construction equipment)

Trailer (trailer or other towed equipment)

Towed artillery

Tank (any size tank or tracked self-propelled gun)

APC (any armored personnel carrier)

Construction T (tracked construction equipment, bulldozers, cranes, etc.)

- - - - - - - - - - - - - - - - -

Firing Trench (trench with firing bays)

Trench (commo trenches)

Foxholes (concentration of 10 or more)

Caves (concentration of 10 or more)

Bldg (building, hut, tent, etc.)

OP (observation post)

Wire (any tactical wire)

Mines (any minefields)

AT (obstacle which was constructed only as an anti-tank obstacle)

Pill boxes

Electronic (radio, radar, etc.)

Airfield

B-3. SAMPLE ADDITIONAL INFORMATION SHEET

Feedback Presented to Groups A and B

Man No. _____

From previous experience, it has been determined that an interpreter's statement of confidence is very important to a commander in evaluating the accuracy of the identifications. Performance measures T-14, T-4, and T-22 have been scored and analyzed. The accuracy of your confidence statements for these measures has been determined by comparing your stated confidence in your identifications of all targets with the percentage of targets correctly identified. The table below summarizes the accuracy of your confidence statements for these three measures. This table indicates how accurate your confidence statements were, whether they were overestimates or underestimates, i.e., whether you were overconfident or lacked confidence in your identifications. Read this table carefully before starting your next performance measure. Try to improve your confidence estimations--90% of the identifications you make with a confidence of "90" should be correct, 80% of the identifications you make with a confidence of "80" should be correct, and so on.

ACCURACY OF CONFIDENCE STATEMENTS FOR PERFORMANCE MEASURES T-14, T-4, T-22

Confidence Intervals	No. Targets Identified	% Targets Correctly Identified	% Confidence Statements	
			Over Estimates	Under Estimates
81-100	9	55	35	
61-80	7	14	56	
60 and below	18	6	24	

B-4. SAMPLE CONFIDENCE ESTIMATE SUMMARY SHEET

Additional Feedback Presented to Group C Only

ACCURACY OF CONFIDENCE ESTIMATES FOR PERFORMANCE MEASURES, T-8 AND T-10							
	Accuracy of Prior Classes				Your Performance		
Confidence Estimate	% Targets Correctly Identified	Estimate		Confidence Estimate	% Targets Correctly Identified	Estimate	
		% Over	% Under			% Over	% Under
100	85	15		100	100	0	0
80-99	76	14		80-99	67	23	
60-79	58	12		60-79	100		30
40-59	38	12		40-59	13	37	
Under 40	25		5	Under 40	25		5

The above table presents accuracy of confidence estimate data from several previous classes on Performance Measures T-8 and -10. Note that the amount of overconfidence for these tests is less than that for T-4, -14, and -22, but there is still room for improvement.

Your performance on T-8 and -10 shows that you are still a little OVERCONFIDENT in your expressions of confidence in your identifications, with the exception of estimates in the 60 to 79 range, where you appear to be a little cautious.

Using this information as a guide, on the next series of Performance Measures, try to adjust your confidence estimates so that they are more accurate. Check each identification carefully before expressing your confidence in the identification. Remember, a target identified with a confidence estimate of 100 should be correctly identified 100 percent of the times, while a target identified with a confidence estimate of 50 should be correctly identified ONLY 50 percent of the times.

APPENDIX C. ANALYSIS OF VARIANCE TABLES

Table C-1

MEAN INACCURACY INDEX

Source	df	MS	F	P
Feedback Methods (M)	3	1154.3	3.78	.05
Subjects = e_1	56	305.1		
Test Series (T)	1	946.4	4.90	.05
M x T	3	12.9	.07	NS
Subjects x T = e_2	56	193.1		

Table C-2

ACCURACY OF IDENTIFICATION

Source	df	MS	F	P
Feedback Methods (M)	3	522.8	.96	NS
Subjects = e_1	56	542.2		
Test Series (T)	1	1833.0	5.99	.05
M x T	3	77.5	.25	
Subjects x T = e_2	56	305.8		

Table C-3

COMPLETENESS OF IDENTIFICATION

Source	df	MS	F	P
Feedback Methods (M)	3	133.1	.85	NS
Subjects = e_1	56	157.3		
Test Series (T)	1	4013.6	30.2	.001
M x T	3	32.9	.25	NS
Subjects x T = e_2	56	132.8		

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) U. S. Army Behavioral Science Research Laboratory, OCRD, Washington, D. C.	2a REPORT SECURITY CLASSIFICATION Unclassified 2b GROUP

3 REPORT TITLE

Impact of Feedback on Accuracy of Confidence Levels Assigned by Interpreters

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

5 AUTHOR(S) (Last name, first name, initial)

Thomas, James A. and Sadacca, Robert

6 REPORT DATE	7a TOTAL NO OF PAGES	7b NO OF REFS
June 1967	29	3

8a CONTRACT OR GRANT NO. b. PROJECT NO. DA R&D Proj. No. 2J620901A721 c. Man-Computer Functions d. b-00	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Research Note 187 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 AVAILABILITY/LIMITATION NOTICES

Qualified requestors may obtain copies of this report directly from DDC. Available, for sale to the public, from the Clearinghouse for Federal Scientific and Technical Information, Department of Commerce, Springfield, Virginia 22151.

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY
	Asst Chief of Staff for Intell, Wash., D.C. Office, Chief of Rsch & Dev, Wash., D. C.

13 ABSTRACT One major objective of the MAN-COMPUTER FUNCTIONS Task is the development of techniques whereby the computer can increase the effectiveness of interpretation by supplementing decision processes of the interpreter, performing his routine calculations, and evaluating accuracy and completeness of interpretations. The present study dealt with the utility of feedback presented under simulated computerized conditions in improving the performance of interpreters in judging the value of their own identifications. Three experimental groups and one control group examined three sets of imagery in the same order, one set on each of three successive days. Confidence feedback based on the previous day's performance was provided each of the experimental groups by a different technique. Interpreters in the control group received no feedback. Levels of confidence in target identifications, stated by the interpreters, were then compared with the scored accuracy of their reports. Results supported previous findings that interpreters do not as a rule make dependable evaluations of their identifications. Confidence ratings made by interpreters in the high performance subgroup were generally more accurate and complete than those made in the low performance subgroup. Two feedback techniques in which the interpreters were given only data on previous rating performance--their own and their own plus that of other classes--resulted in somewhat more accurate expressions of confidence than did the technique in which interpreters were given their own corrected reports and the imagery they had previously interpreted. Confidence ratings reported by the interpreter group receiving no feedback were the least precise. It was concluded that interpreters' confidence ratings can be improved by practice in applying a

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
*Image systems--work methods						
Image interpretation						
Intra-system information--transfer						
Computer-based image interpretation systems						
Laboratory facilities						
*Feedback techniques						
*Image interpretation performance measures						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13 ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

DD Form 1473

13. ABSTRACT continued

knowledge-of-results frame of reference. Findings suggest, however, that more than two practice sessions are needed for the interpreter to reach an operationally useful level of accuracy in evaluating the information he provides.